



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

THE SCALE ILLUSION

C. H. WARD

The Taft School, Watertown, Connecticut

Suppose that a recruiting sergeant had strayed into a convention of professors who were discussing the question, "What are the qualifications of a private soldier?" And suppose that one of the savants had proposed a physique scale, a row of twenty men ranging in military fitness from the zero of a consumptive to the 95 per cent of an all-America football player. The rejecting of volunteers is no longer to be a matter of the caprices of examining physicians, but is to depend upon a consensus of the opinions of all sorts of experts; there are no longer to be examinations of eyes, teeth, lungs, and heart, but instead a mass-decision as to whether a candidate ranks nearest in fitness to specimen No. 13 or No. 17. What would the ignorant sergeant say to this discussion? He could say nothing. He would be dazed at finding reverend men seriously occupied over such a proposal, would begin to doubt his own reason, would hasten out into the fresh air.

A mere preparatory teacher is dazed when he hears serious consideration of a scale for grading composition. He is a recruiting sergeant, ignorant of academic possibilities, self-distrustful, respectful toward his betters; he does not contend that grading by a scale is scholastically impossible; he is simply unable to apprehend how such a proposal is related to real life. If he were required to argue his distrust before Professors Hillegas and Thorndike, he would be embarrassed, quite unequal to the task. But if he were asked to describe his feelings about grading by scale, to give this description merely as one bit of evidence, he would say, "Anything for the good of the service."

The preparatory teacher of composition is a queer animal. Hardly anything is known about him. High-school teachers suppose that he is a slave to entrance requirements; colleges consider him a crafty imparter of counterfeit knowledge. But everybody

knows that he has one characteristic: he gets his living by adapting himself to an objective standard of theme-grading. If he is to earn a salary, he must put aside all preconceptions of composition values, must find out by hook and crook what kind of candidate flunks, must learn that literary genius fails and prosy carefulness gets a fair mark, must whip the genius into some command of spelling, and spur the prosy pupil into some aptitude for progress to a snappy conclusion. He must go much farther than this. He must be a Sherlock Holmes at inferring what kind of errors count most, so that his greatest effort will be directed at the most vital matters. I say "inferring," for until 1913 no information was ever given him about how college readers graded; they marked in secret, obliging the teacher to play blindman's buff as best he might. In this game he availed himself greedily of every scrap of evidence: he learned that last year's papers were conditioned for four misspelled words; he guessed at marks on papers before they went to examiners; he very casually inquired of a college instructor whether "writting" was worse than "village." In the course of ten years he was able to formulate a scheme of grading that corresponded very closely with college results: tabulations for five consecutive years showed that 96 per cent of the boys who had a school mark of 60 or better got 60 or better on the college examinations; boys who had a school mark of 55 were almost sure to get below 60 on the college examinations. His arithmetical correspondences matched up in another and more convincing way when the school averages for 1915-16 were compared with the Board marks for 1916. The examinations were taken by 52 boys; 1 got just his school mark, 64; 23 had been rated higher by the school than by the Board, 28 rated lower by the school; the average of the overratings by the school was 8.30, the average of the underratings 7.96; the maximum overratings were 29, 21, 16, 15, the maximum underratings 25, 24, 19, 15; there were 28 school marks within 6 points of the Board ratings. Closer than this the school hardly hopes to get, because the variations are not greater than the normal differences between its own examinations and recitation averages.

All this kind of tabulation may be miserable business. It is not paraded here as useful in itself. It is simply an experiment

in objective standards, a proof that composition-grading has been somehow standardized by college readers, that their standard is almost invariable, and that it can be closely approximated by teachers who have never been told what it is. If the world is looking for a standard, why should it not at least start with one that has been evolved in the course of twenty-five years of the practice of experts? Grant that their knowledge of psychology is slight and their sense of literary values badly warped; they are at least practical persons, adept in the actual business of judging the writing of youth.

Is it contended that they are too arithmetical, too insistent upon mere accuracy? Perhaps they are. But the preparatory animal never raises a protest of that sort, never presumes to question the good sense of those who admit to college, any more than other animals lift their voices against the harshness of the world in which they find themselves. They use their claws and noses to get a living; so does he. He digs hard and sniffs at every breeze that may tell him where prey can be found. If he can hunt up enough college-entrance food, he lives; if not, he dies. But he dies without whimpering at the unfairness of being obliged to lurk and prowl for sustenance.

And just as a weasel cannot conceive of a dining-room where carnivorous mammals have to refrain from overeating, so the preparatory animal cannot imagine that aesthetic region in which themes are graded by a general impression. It is literally true that until the Hillegas scale was mooted he had never realized that there existed such a process as impressionistic grading. Though he is told of it, he cannot yet conceive it. When he reads the now familiar story of how theme No. 14 was marked 20 by A and 95 by B, and theme No. 15 was marked 90 by A and 25 by B, he has no experience that makes it comprehensible. For he knows that two preparatory teachers grading according to their school standards for a given class would never show such wild inversions of judgment. His assurance is not immodesty; he has no more pride than a weasel has about his claws. When he hears about "doing away with impressions," he knows that the proposed scale is merely a substitution of one kind of subjectivity for another. Any measure of literary value is impressionistic; any measure of

literary value and mechanical value at the same time is a phantom; and if the proposed scale is a measure of mechanics only, it is the roughest kind of guesswork.

Nor is he moved by accounts of how the teachers at Oldburg High learned with practice to reduce their variations immensely. That means only that they grew accustomed to rating errors small and big, and to feeling their way toward a less careless guess. "But isn't this better than impressionism?" you ask him. And he replies, "Infinitely better. It is to some extent objective. But if you want objectivity, why do you toy with this theoretical makeshift? Why don't you get a real scale?"

If he has to explain, he will revert to that allegory of the recruiting sergeant, which by its concreteness furnishes real illustration. Consider first how utterly unfair some of the requirements for enlistment are. A man is refused because he is only sixty-three inches tall, though everybody knows that some sixty-inch men would fight like demons; a man is refused because he is forty years old, yet millions of soldiers have been valiant after they were fifty. Hence we can fancy how some intellectual civilian might argue that these requirements are illogical, that a scale of twenty graduated physiques would be a more scientific measure of all-round fitness. He demonstrates that sergeants can pass judgment by means of such a scale, and exclaims, "So, you see, it works."

"But it would work just the same," says the unmoved officer, "if it wasn't there. No man of experience is helped by seeing a lot of human beings in a row. I can tell at a glance whether a fellow is fit. And I don't get any 'consensus' or 'composite,' or whatever you call it. I'm seeing whether each separate requirement is there. There isn't any such thing, sir, as a 'consensus of competent opinion' that you could get to make up your scale by. Nobody that knows anything about picking men would ever have use for such a scale. A good man has got to be good in several respects—in each one of those respects. And there's nothing hard and fast about our method, either. In peace times if lots of men are offering, we can make every requirement strict and high. If the need for soldiers increases, we can lower one requirement, or all requirements, or cut out one, or cut out two and lower the rest—we can fit any demand. But we've got to work with those

units, sir. If any medical examiner thinks he works with a whole man at once, he's fooling himself."

"Suppose we grant all you say," rejoins the philosopher. "The most monstrous fault of your system has not been touched upon. You are applying merely mechanical measurements. The chief attributes of soldiership are, after all, quite different—bravery, initiative, resourcefulness, endurance. You take no stock of these."

"Well, sir, it sounds funny to say so, but the fact is we know that in sound bodies we are sure to get the average of bravery and endurance. And we can't get more than an average except in picked troops. And as for initiative and all that—why, you're talking about officers now. See it?"

The sergeant's ideas about men apply to the intangibilities of theme-grading. To insist upon decent spelling, idiom, punctuation, and thought-sequence, to condition a theme for falling short in any one particular, to insist upon examining each of the requisites, to have an arbitrary standard of conditioning for six misspellings or two sentence-errors, to reject a scale as a fantastic theory, to have no schedule of clear thinking, charm, effectiveness—to rate themes in this way may prove that a teacher is soldierish. But his rude method is essentially the one that has always been evolved when a body of instructors has had to grade composition uniformly.

The principle is always this: Subtract from a perfect mark according to an arbitrary scheme of errors. Until the University of Wisconsin published in 1913 its *Requirements for Admission to the Freshman English Course* there was not, so far as the writer knows, any available information as to an established scheme of errors. This bulletin performed a great and novel service by premising in these daring terms: "Students whose writing is devoid of interest, originality, or any other literary merit are qualified if their writing is satisfactory as to rudiments; and students who possess literary skill are not qualified if they are seriously deficient in the rudiments." The body of instructions is composed of a series of "don't's," rated according to seriousness—another novelty as startling as it was sensible. Five faults are "extremely serious," thirty-six are "less serious," but "more serious than the other forty-six." It is impossible that the

Wisconsin ranking is generally acceptable, for it classes all the misspellings as less serious than such supposedly bad diction as "burglarize" and "near-by." This is topsy-turveydom. But the principle is valid, is indispensable: errors vary infinitely in importance.

Working without any such clue, but guessing his way gradually, the preparatory teacher has for twenty years been setting up his own scheme of errors. The process has been long and complicated; for the most part it has consisted of mere instinctive decisions, not of deductions and tabulations. The result is too bulky for presentation here, and has not been formulated; but in outline it is like this: There is a "unit error"—omitting an apostrophe or a comma before "but," such a misspelling as "beleive" or "accidently," failing to indent, failing to paragraph for dialogue. In the lowest class these unit errors may count four each on short themes as soon as the class has been drilled in each particular; for the next class they may count, on longer themes, five each at the beginning of the year, and as the year goes on new items are being added to the list. In proportion as any matter—say "varius" or "dissappear"—is typical and has received special emphasis it is counted double, and a few specialties are counted triple. The comma fault counts five units. Thus in the case of a pupil who is nearly ready to take college examinations 20 points may be deducted from a theme-grade for "strenuos," nothing for "strenuous," 10 points for "dont," 20 for "dident," 35 for "to much," and 50 for a comma fault. Poor paragraphing or poor order of ideas will be guessed at as amounting to two or three—seldom to four—unit errors. The "higher matters" almost take care of themselves, because—note well—a pupil who qualifies in small matters is never egregiously incoherent.

Correctness in spelling and punctuation does not constitute a good theme. By imputing that meaning to the preceding paragraph a reader can beget in himself a feeling of contempt for this whole thesis. It has no such meaning. But the fact of experience is—we hope this can be put emphatically—that *the sort of mind* that can learn careful accuracy is almost infallibly *the sort of mind* that can be orderly and effective, and vice versa. It is an ancient saying and worthy of all acceptance that he who is competent in

the little things is competent also in the big ones. Just as the sergeant need not have any rating of bravery and initiative, so the composition-grader need not have any scheme of charm and effectiveness. The ordinary modicum of those personal qualities will be present; more than the modicum is not to be had.

In November, 1915, four Freshman instructors of the University of Illinois presented in the *English Journal* a detailed statement of their system of subtraction—the first ever seen by the writer. They bring down a grade to below passing for any of the following errors: 3 misspelled words; 2 straggling sentences; 1 comma fault; 2 grammatical errors; a noticeable number of improprieties or barbarisms; a marked lack of unity; a marked lack of coherence.

In one way this differs from the Wisconsin scheme: improprieties seem to count less than misspellings. It is much more nearly in accord with preparatory grading. A grade of D may be obtained for mere freedom from mechanical errors, but for higher marks there must be at least one of the "positive virtues." Illinois concludes as Wisconsin begins: "Originality without literacy, without orderly and accurate expression, cannot in any place or under any circumstances be considered a meritorious quality in a college course in English composition." The whole article shows how far the authors are from any faith in a Hillegas scale. By adaptation to their peculiar needs they have arrived at just the kind of system that is reached by all practical strivers after an objective standard—a system of equating errors and subtracting from a perfect mark.

What does the College Board do? Its method would have more authority than any other, because its readers are all experienced, come from a great variety of schools and colleges, are organized into a homogeneous corps, maintain year after year a steady uniformity. If there is any such thing as an impersonal criterion that represents a wide consensus of opinion, it must exist in the committee of readers of English papers. Their judgment is sometimes called severe, but has never been capricious. Any college that admits only for a Board mark of 60 will not find much illiteracy among its Freshmen, will not have to do the elementary work necessary for freshmen admitted by certificate; a college that admits on 40 will find its students unfit for decent

scholarship; those who can get only 55 are doubtful college material. The point is that the Board has a standard that means something, that can be interpreted everywhere, that is stable. Nor is it new. The writer adapted his teaching to Yale examiners for seventeen years; when the change was made to the Board in 1916, he found the new criterion just like the old one. How the Board grades are arrived at has never been divulged. In secret they have sat, and a secret their code remains. But now the curtain is partly drawn. Their *Document No. 80*, published last November, though it gives no details, reveals clearly enough that the method is a subtraction for shortcomings and errors. Four themes are reprinted. One, we are told, is irrelevant and haphazard in execution, and therefore "the writer failed badly." (We note incidentally that it is speckled with eighteen petty errors.) The second theme, "in spite of force and interest," is called poor because of bad planning and a few mistakes. The third is "well assembled," but "in spite of the merits of the writing" the book containing it fell a few points below sixty, because the writer is "more than usually careless in punctuation and spelling." In the fourth there is "thoroughly logical arrangement and relative freedom from careless errors." (There is no error that is typical except two omissions of unimportant commas.) The preparatory experience is corroborated: a theme that is free from typical petty errors is almost certain to be a decent theme in larger matters.

Document No. 80 has almost nothing to say about the improprieties that bulk so large in the Wisconsin bulletin. It announces with weighty brevity: "The first requisite in composition is correctness in mechanical details." All its advice is sound and helpful. It will be a godsend to many a teacher who has been befogged in the vague notions that somehow school themes are matters of expressing personality with "fluency," and that they are best evaluated by reading a literary vernier.

Why has the Board not been more specific? The answer could doubtless be furnished by the curator of the Bronx Park, who once had a rare Arctic hawk that would eat only living birds, and who had to keep his cruel feeding a secret for fear of the S.P.C.A. A conscientious public always knows more about zoölogical justice

than can be conceived by a heartless naturalist. So we may suppose that a complete display of the heartless procedure of Board readers of English would encounter a wrathful visitation from a Society for the Prevention of Strictness with Candidates. There are many generous and well-meaning souls who are pleased enough with our fierce hawk of a committee while they see only results, but who might raise a rumpus and seek to destroy that committee if they knew the details of how it lived. Otherwise we can see no reason why the schools should not know more precisely how the Board grades. (Even in *Document No. 80* we are not told the grades of themes, but only the mark on the whole paper of which the theme was a part.) Such detailed information might have a tendency to make mechanical teachers more mechanical—a sad result, but a one-per-cent sadness. The 99 per cent of benefit would be in spurring up all good teachers to put more emphasis on carefulness. Our training in composition, now so inadequate and slovenly, would be most wholesomely directed to true education; our business men would have less occasion for cursing the schools, and our universities could be freed from some of their present seventh-grade labors.

That cursing and those labors will continue—yes, are bound to increase—if anything like the scale illusion makes progress. What practical men and women use a scale? Probably dozens of hopeful experimenters think well of it. It has some vogue in a form known as the Harvard-Newton scale. It may have trued up theme-reading in many a school that used to get along with no criterion at all. It may be used in some paradisaic regions where no attention need be paid to specific rudimentary errors, where the desideratum is to save time or to get a mere record of the height above some absolute zero. But theme-reading in real life always has a purpose. In the grades we must be leading the child to see what simple sentences are and how they should follow one another in some order. In the grades at Lawrence, Massachusetts, Superintendent Sheridan is conducting a sentence campaign that may in time set American universities free from their slavery to comma faults; what mention would he make of a scale in his emancipation proclamation? In high schools we should be leading the pupils

to see what a complex sentence is and what a paragraph is. In what way could a scale give light to pupils or remove the honest teacher's duty of thorough reading? Could the Board readers justify themselves for admitting to college on the basis of "my general feeling with reference to specimen No. 17"? Even if they should adopt scale-grading, the fact of their mental process would be this: "About how many, and how serious, errors did I note? Down to what specimen of the scale would these drag the composition?" That must always be the fact about attaining accurate results by the use of a scale—subtraction for errors. The presence of a scale merely substitutes "about" for "exactly." Have college students got above the region of rudiments? Williams has had to devote the opening weeks of its year of composition to drill in simple punctuation. Two years ago Mr. Hersey of Harvard reported that of the students in the Graduate School of Business Administration (all college graduates) only 21 per cent received passing marks when their reports were graded as composition. Did he recommend a scale? No, he issued explicit "Directions for Writing." In real life the purpose of theme-grading is to show students how they go wrong and how they may improve. Does any body of college instructors hope to improve style by adopting a scale of specimens?

Has France developed any scale-reading device? Professor Brown declares in *How the French Boy Learns to Write*: "It seems to be clearly understood by French teachers of the mother tongue that the grading of themes implies long hours of patient labor. I found no teacher who professed to believe that anybody could discover a 'royal incline' that would save one from the annoyance of serious effort. The great heavy burden of theme-reading is regarded by teachers as one of the inevitable but fruitful duties of their profession."

"Fruitful" is the word. It means that theme-reading is to produce results in the pupil's intellect. A system of subtraction, that shows him exactly how he can improve, bears fruit. A system that shows him only his height above an absolute zero can no more produce a harvest than a thermometer can bring forth figs.